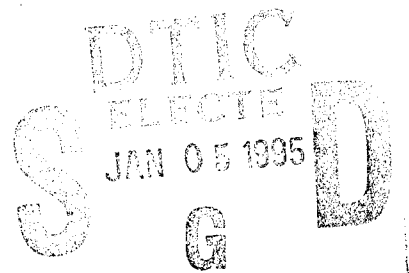


## Can There Be Reliability without "Reliability?"

Robert J. Mitlevy



19950103 083

This research was sponsored in part by the  
Cognitive Science Program  
Cognitive and Neural Sciences Division  
Office of Naval Research, under  
Contract No. N00014-91-J-4101  
R&T 4421573-01

Robert J. Mitlevy, Principal Investigator



Educational Testing Service  
Princeton, NJ

October 1994

Reproduction in whole or in part is permitted  
for any purpose of the United States  
Government.

Approved for public release; distribution  
unlimited.

DTIC QUALITY INSPECTED 3

REPORT DOCUMENTATION PAGE			Form Approved OMB No. 0704-0188	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.				
1. AGENCY USE ONLY (Leave blank)		2. REPORT DATE October 1994		3. REPORT TYPE AND DATES COVERED Final
4. TITLE AND SUBTITLE Can There Be Reliability without "Reliability?"			5. FUNDING NUMBERS G. N00014-91-J-4101 PE. 61153N PR. RR 04204 TA. RR 04204-01 WU. R&T 4421573-01	
6. AUTHOR(S) Robert J. Mislevy				
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Educational Testing Service Rosedale Road 03-T Princeton, NJ 08541			8. PERFORMING ORGANIZATION REPORT NUMBER RM-94-18-ONR	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Cognitive Sciences Code 1142CS Office of Naval Research Arlington, VA 22217-5000			10. SPONSORING/MONITORING AGENCY REPORT NUMBER N/A	
11. SUPPLEMENTARY NOTES None				
12a. DISTRIBUTION/AVAILABILITY STATEMENT Unclassified			12b. DISTRIBUTION CODE N/A	
13. ABSTRACT (Maximum 200 words) A recent article by Pamela Moss asks the title question, "Can there be validity without reliability?" If by reliability we mean only KR-20 coefficients or inter-rater correlations, the answer is yes. Sometimes these particular indices for evaluating evidence suit the problem we encounter; sometimes they don't. If by reliability we mean credibility of evidence, where credibility is defined as appropriate to the intended inference, the answer is no, we cannot have validity without reliability. Because "validity" encompasses the process of reasoning as well as the data, uncritically accepting observations as strong evidence, when they may be incorrect, misleading, unrepresentative, or fraudulent, may lead coincidentally to correct conclusions but not to valid ones. This paper discusses and illustrates a broader conception of "reliability" in educational assessment, to ground a deeper understanding of the issues raised by Professor Moss's question.				
14. SUBJECT TERMS Educational assessment, hermeneutics, reliability, validity.			15. NUMBER OF PAGES 23	
			16. PRICE CODE N/A	
17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT SAR	

# Can There Be Reliability without "Reliability?"

Robert J. Mislevy  
Educational Testing Service

October, 1994

Accession For		
NTIS	CRA&I	<input checked="checked" type="checkbox"/>
DTIC	TAB	<input type="checkbox"/>
Unannounced		<input type="checkbox"/>
Justification		
By		
Distribution /		
Availability Codes		
Dist	Avail and/or Special	
A-1		

This paper is a response to Pamela Moss's (1994) *Educational Researcher* article, "Can there be validity without reliability?" It was supported by Contract No. N00014-91-J-4101, R&T 4421573-01, from the Cognitive Science Program, Cognitive and Neural Sciences Division, Office of Naval Research, and by the National Center for Research on Evaluation, Standards, and Student Testing (CRESST), Educational Research and Development Program, cooperative agreement number R117G10027 and CFDA catalog number 84.117G, administered by the Office of Educational Research and Improvement, U.S. Department of Education. I am grateful to Drew Gitomer, Charlie Lewis, and Howard Wainer for discussions and helpful comments on an earlier draft.



## Can There Be Reliability without "Reliability?"

### Abstract

A recent article by Pamela Moss asks the title question, "Can there be validity without reliability?" If by reliability we mean only KR-20 coefficients or inter-rater correlations, the answer is yes. Sometimes these particular indices for evaluating evidence suit the problem we encounter; sometimes they don't. If by reliability we mean credibility of evidence, where credibility is defined as appropriate to the intended inference, the answer is no, we cannot have validity without reliability. Because "validity" encompasses the process of reasoning as well as the data, uncritically accepting observations as strong evidence, when they may be incorrect, misleading, unrepresentative, or fraudulent, may lead coincidentally to *correct* conclusions but not to *valid* ones. This paper discusses and illustrates a broader conception of "reliability" in educational assessment, to ground a deeper understanding of the issues raised by Professor Moss's question.

Key words: Educational assessment, hermeneutics, reliability, validity.

## Introduction

"Can there be validity without reliability?" asks Pamela Moss (1994) in her article of the same name in the *Educational Researcher*. Yes, Professor Moss answers. She proposes a hermeneutic approach to educational assessment—the validity of which, she argues, does not depend on standard test theory indicators of reliability such as KR-20 coefficients and inter-rater correlations. I agree that it is possible have validity without reliability, if by "reliability" we refer only to these particular indices and others like them, but this is far too narrow a conception of "reliability." More broadly construed, however, reliability concerns the credibility and the limitations of the information from which we wish to draw inferences. If we fail to address this concern *in an appropriate manner*, we fail to establish the validity of those inferences. This paper discusses and illustrates a broader conception of "reliability" in educational assessment, to ground a deeper understanding of the issues raised by Professor Moss's question. (See Mislevy, 1994, for a more comprehensive discussion.)

That reliability be examined "in an appropriate manner" is key, because KR-20s and inter-rater correlations characterize the credibility of certain kinds of data we employ for certain kinds of inferences in educational assessment, but not others. I applaud Professor Moss's use of a hermeneutic perspective to gain insights into questions of educational assessment, because, as Goethe wrote in *Sprüche in Prosa*, "He who is ignorant of foreign languages knows not his own." Exploring how other fields deal with evidence and inference can indeed help us disentangle the commingled concepts from statistics, psychology, and measurement that constitute test theory as we usually think about it—to distinguish *how we are reasoning* from *what we are reasoning*

*about*—to better prepare ourselves to tackle problems of how to characterize students' learning beyond scores on standardized tests, how to evoke and interpret evidence to this end, and how to establish the weight and coverage of data as evidence for conjectures and decisions framed in these terms. Physical measurement has long been a source of concepts and techniques for educational assessment (e.g., Rasch, 1960/1980). In addition to the hermeneutic tradition, we can also gain insights from fields such as medicine, history, and jurisprudence<sup>1</sup> (Schum, 1987). Seeing how "reliability problems" arise and how they are dealt with in these fields helps us understand their appearance in our own.

### What is Reliability?

We can think of increasingly general senses of the term "reliability" as it relates to educational assessment:

True-score reliability (Gulliksen, 1950). The classical reliability coefficient  $\rho$  assumes repeatable observations comprised of an examinee's "true score" and a random "measurement error."  $\rho$  is the proportion of variance in a particular population of examinees' observed scores attributable to the variance of their true-scores. The data are equally-valued responses to interchangeable tasks, constituting a source of potentially *collaborating* evidence, or more evidence of the same kind about a given inference.  $\rho$  does in fact gauge observed scores' weight of evidence—for the inference of

---

<sup>1</sup> In *Analysis of evidence*, Anderson and Twining (1991) use analogies from educational testing to help law students learn distinctions among rules, criteria, standards for evaluating evidence.

lining up people from this particular population along the true-score scale. It does in fact bound "validity"—for the inference of predicting a variable related linearly to true score, with "validity" defined as the correlation between the scores and predicted variables in this particular population. But even under classical test theory, rho need not convey the evidential value of scores for other inferences—for example, the magnitude of change in true score from pretest to posttest, or whether a student's true score is above a specified cutoff value.

Reproducibility. We can extend "reliability" beyond this specific and population-bound inference, yet retain grounding in the consistency of exchangeable (equally-informative and equally-valued) independent sources. Experimenters attempting to reproduce Pons and Fleishmann's purported cold fusion results faced reliability concerns in this sense: "The way to circumvent this skittishness [of BF<sub>3</sub> neutron counters] was to use two counters, or even five or six, and only pay attention to those events in which all the detectors fired simultaneously" (Taubes, 1993, p. 450). Before detailing investigations of witnesses to the Kennedy assassination, Gerald Posner (1993, p. 236) summarized an overarching pattern:

How many shots were fired at Dealey Plaza? ... Estimates at the scene ranged from one to eight. However, on this issue, there was more agreement than on any other postassassination matter. Of the nearly two hundred witnesses ... over 88 percent heard three shots. ...

Although almost every conspiracy theory proposes that more than one assassin relies on there having been four or more shots, the writers seldom disclose that fewer than one in twenty witnesses heard that many.



In educational measurement, proportions of agreement among raters, decision-consistency coefficients, and generalizability coefficients (Cronbach *et al.*, 1972) reflect this sense of reliability. These indices characterize the weight of evidence for inferences within the true-score test theory paradigm that are not addressed by rho. They can be useful even if one doesn't literally believe observations are exchangeable. The "Concentration" section of each Advanced Placement Studio Art portfolio is rated by two judges independently, and only portfolios that provoke excessive differences are probed further. If ensuing discussion reveals one judge differed because of special knowledge of, say, glazing techniques, this information impacts the deliberation. The exchangeability framework provides indices of similarity among judges' evaluations, but just as importantly, it highlights particulars where exchangeability is not a plausible approximation, to direct attention and expertise where they are most needed (Myford & Mislevy, in press).

Differential likelihood. "A datum becomes evidence in some analytic problem when its *relevance* to one or more hypotheses being considered is established. ...[E]vidence is relevant on some hypothesis if it either increases or decreases the likeliness of the hypothesis" (Schum, 1987, p. 16). Under probability-based reasoning, the relative likelihood of an observation under alternative "true states" is the weight of evidence it provides for each; "reliable" observations make sharp distinctions among the possibilities.<sup>2</sup> The

---

<sup>2</sup> Agreeing *too much* on key points, along with agreeing *too little* on tangential issues, lowers the credibility of suspected collaborators in a criminal investigation; this pattern is likely under the hypothesis of a rehearsed alibi. Reproducibility does not equal to credibility, since a

empirical consistency discussed above is one way to ground likelihoods; we take a BF<sub>3</sub> burst with a grain of salt once we know a bump is as likely to cause one as an actual neutron. Theoretical and subjective considerations can also provide information about relative likelihoods. The MUNIN neuromuscular disease diagnostic system uses conditional probabilities for test results and symptoms given disease states, which are based on clinical experience and physiological theory (Andreassen, Jensen, & Olesen, 1990). Failure Analysis Associates' "probability cones" (Figure 1) for sources of shots in the Kennedy assassination extend uncertainties in positions and angles backwards from the points of impact (Posner, 1993, p. 476).

[Figure 1]

We use similar reasoning to convey our uncertainty about a student's proficiency under an item response theory model, or her stage of proportional reasoning under a latent class model. We obtain in these cases numerical assessments of the evidential value (read "reliability") of the data—but only if, perhaps after considerable effort, we can arrange circumstances in which our data, our model, and our intentions cohere (Wright & Stone, 1979). Less formally, a tutor constructs a model for a student's understanding, probing "What organization does the student have in mind so that his actions seem, to him, to form a coherent pattern?" (Thompson, 1982). "Reliable" data allow the tutor to identify a perspective from which the student's pattern of actions make sense, but are unlikely from relevant alternative perspectives.

---

conspirator can repeat a lie 100 times. Data "too good to be true" toppled Cyril Burt (Kamin, 1974).

This is not "reliability" in the sense of accumulating *collaborating* evidence, as in classical test theory, but in the sense of *converging* evidence—accumulating evidence of different types that support the same inference. A mass of data is more reliable in this sense as more aspects support a given inference and fewer aspects conflict or contradict it; it is less reliable if when it is internally inconsistent or equivocal, or if we realize that securing additional information would cause us to revise our beliefs substantially. Such considerations characterize the reliability of the evidence supporting a legal case, and jurists and statisticians have explored the means by which, and the extent to which, they can be expressed in terms of differential likelihoods (e.g., Kadane & Schum, 1992).

Credibility. In common parlance, reliability simply means the extent to which information can be trusted, a concern clearly broader than traditional educational measurement situations. The world constantly confronts us with unrepeatable observations and non-exchangeable sources, which we must interpret as best we can if we have no alternative (there was only one trial of the Kennedy assassination), or learn from to develop more principled ways of gathering and interpreting information (to assess prospects of cold fusion or students' understandings of proportional reasoning).

When sources are not exchangeable, we must unravel secondary sources of information about *their* credibilities:

- Not all cold fusion experiments are created equal; those with better controls and more reliable measuring instruments, or incorporating lessons from earlier experiments, are privileged. Early positive results were traced to experimental mistakes and interpretational errors, in

which questionable data were consistently accepted as evidence of desired outcomes (Taubes, 1993).<sup>3</sup>

- *Lincoln at Gettysburg* (Wills, 1992) is mainly a hermeneutic analysis of what Lincoln *meant* when he presented the Gettysburg Address, but its Appendix I explores what he actually *said*. Five versions in Lincoln's hand and four newspaper transcriptions survive. Unanimity about a phrase suggests he spoke it as such, but for discrepancies Wills must consider such clues as these: The draft Lincoln's secretary claimed he saw Lincoln speak from appears on Executive Mansion letterhead, corroborating eyewitness accounts, but omits key phrases all newspapers report and garbles the transition between pages.

We must often integrate multiple strands of evidence, and "reliability" typically refers to the weight of evidence of a particular strand. Influence diagrams in troubleshooting (e.g., Klempner et al., 1991), medical diagnosis (Andreasson et al., 1990), and legal reasoning (Wigmore, 1937) depict how sources and credibilities of information relate to inferences. Temperature is one strand of evidence in determining whether a child's infection is bacterial or viral (Figure 2). A thermometer reading is direct evidence about temperature, and the "reliability" of the thermometer concerns its credibility about this symptom. The reading is indirect evidence about nature of illness.

---

<sup>3</sup> A joke made the rounds of experimental labs: "Q: Why can't most people get heat, neutrons, and tritium [putative evidence of cold fusion] at the same time? A: It's almost impossible to make that many mistakes at once" (Taubes, 1993, p. 468).

In conjunction with other evidence, even a hand on his forehead—an “unreliable thermometer”—can aid in the diagnosis.

[Figure 2]

In this light, the irony is not that test administrators warn test users against interpreting scores without other sources of information, but that the test users themselves are most prone to reify “traits” such as “IQ” or “writing ability.” The view among contemporary researchers, whose work is beginning to influence the next generation of tests, substantiates the caveat:

The evidence from cognitive psychology suggests that test performances are comprised of complex assemblies of component information-processing actions that are adapted to task requirements during performance... Whatever their practical value as summaries, for selection, classification, certification, or program evaluation, the cognitive psychological view is that such [trait-based] interpretations no longer suffice as scientific explanations of aptitude and achievement constructs. (Snow & Lohman, 1989, p. 317).

## Conclusion

Can we have validity without reliability? If by reliability we mean only KR-20 coefficients or inter-rater correlations, the answer is yes. Sometimes these particular indices for evaluating evidence suit the problem we encounter; sometimes they don't. But when multiple sources of evidence *are* available and they *don't* agree, we'd better have alternative lines of argumentation to establish the weight and relevance of the evidence to the inference being drawn. Sometimes people disagree because they focus on different aspects of a situation from different perspectives, which need to be integrated in a more thoughtful way than averaging. But sometimes people disagree because they are uninformed or biased, because their task is not

clearly specified, or because they are dishonest. We bear the burden of unraveling these possibilities.

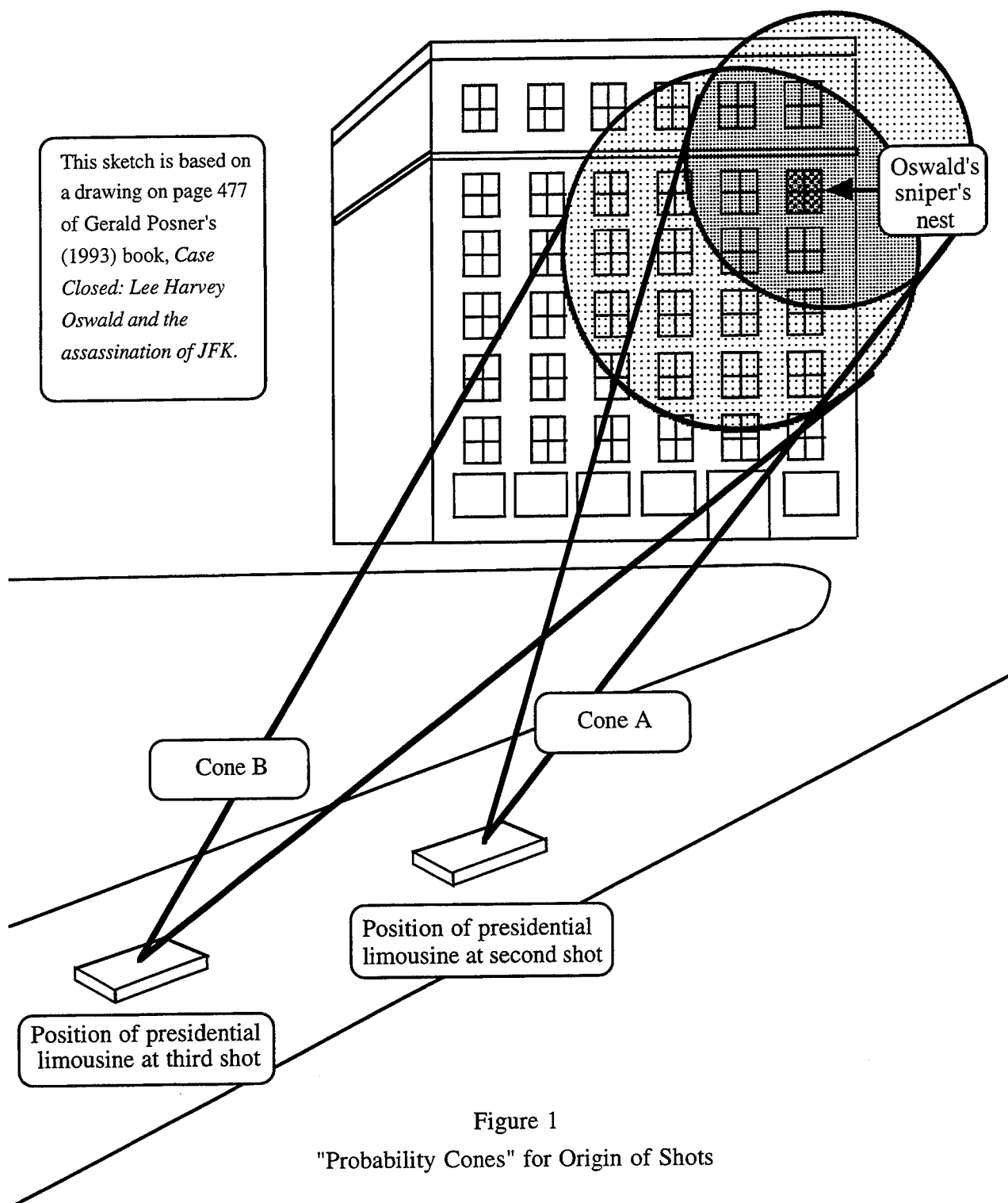
If by reliability we mean credibility of evidence, *where credibility is defined as appropriate to the inference*, the answer is no, we cannot have validity without reliability. Because "validity" encompasses the process of reasoning as well as the data, uncritically accepting observations as strong evidence, when they may be incorrect, misleading, unrepresentative, or fraudulent, may lead coincidentally to *correct* conclusions but not to *valid* ones. Good intentions and plausible theories are not enough to honestly evaluate and subsequently improve our efforts. That familiar tools for establishing the credibility of evidence in educational assessment do not span the full range of inferences does not negate the responsibility to establish the credibility of evidence upon which educational decisions are made. If anything, our task becomes harder rather than easier.

## References

- Anderson, T.J., & Twining, W.L. (1991). *Analysis of evidence*. Boston: Little, Brown, & Co.
- Andreassen, S., Jensen, F.V., & Olesen, K.G. (1990). Medical expert systems based on causal probabilistic networks. Aalborg, Denmark: Institute of Electronic Systems, Aalborg University.
- Cronbach, L.J., Gleser, G.C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements: Theory of generalizability for scores and profiles*. New York: Wiley.
- Gulliksen, H. (1950/1987). *Theory of mental tests*. New York: Wiley. Reprint, Hillsdale, NJ: Erlbaum.
- Kadane, J.B., & Schum, D.A. (1992). Opinions in dispute: the Sacco-Vanzetti case. In J.M. Bernardo, J.O. Berger, A.P. Dawid, & A.F.M. Smith (Eds.), *Bayesian Statistics 4* (pp. 267-287). Oxford, U.K.: Oxford University Press.
- Kamin, L.J. (1974). *The science and politics of IQ*. Potomac, MD: Erlbaum.
- Klempner, G., Kornfeld, A., & Lloyd, B. (1991). The EPRI generator expert monitoring system: Expertise with the GEMS prototype. Presented at the American Power Conference, May, Chicago, IL.
- Mislevy, R.J. (1994). Evidence and inference in educational assessment. *Psychometrika*, 59, 439-483.
- Moss, P. (1994). Can there be validity without reliability? *Educational Researcher*, 23(2), 5-12.
- Myford, C.M., & Mislevy, R.J. (in press). Monitoring and improving a portfolio assessment system. ETS Research Report. Princeton, NJ: Educational Testing Service.

- Posner, G. (1993). *Case closed: Lee Harvey Oswald and the assassination of JFK*. New York: Random House.
- Rasch, G. (1960/1980). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danish Institute for Educational Research/Chicago: University of Chicago Press (reprint).
- Schum, D.A. (1987). *Evidence and inference for the intelligence analyst*. Lanham, Md.: University Press of America.
- Snow, R.E., & Lohman, D.F. (1984). Toward a theory of cognitive aptitude for learning from instruction. *Journal of Educational Psychology*, 76, 347-376.
- Taubes, G. (1993). *Bad science: The short life and weird times of cold fusion*. New York: Random House.
- Thompson, P.W. (1982). Were lions to speak, we wouldn't understand. *Journal of Mathematical Behavior*, 3, 147-165.
- Wigmore, J.H. (1937). *The science of judicial proof* (3<sup>rd</sup> Ed.). Boston: Little, Brown, & Co.
- Wright, B.D., & Stone, M. (1979). *Best Test Design*. Chicago: MESA Press.





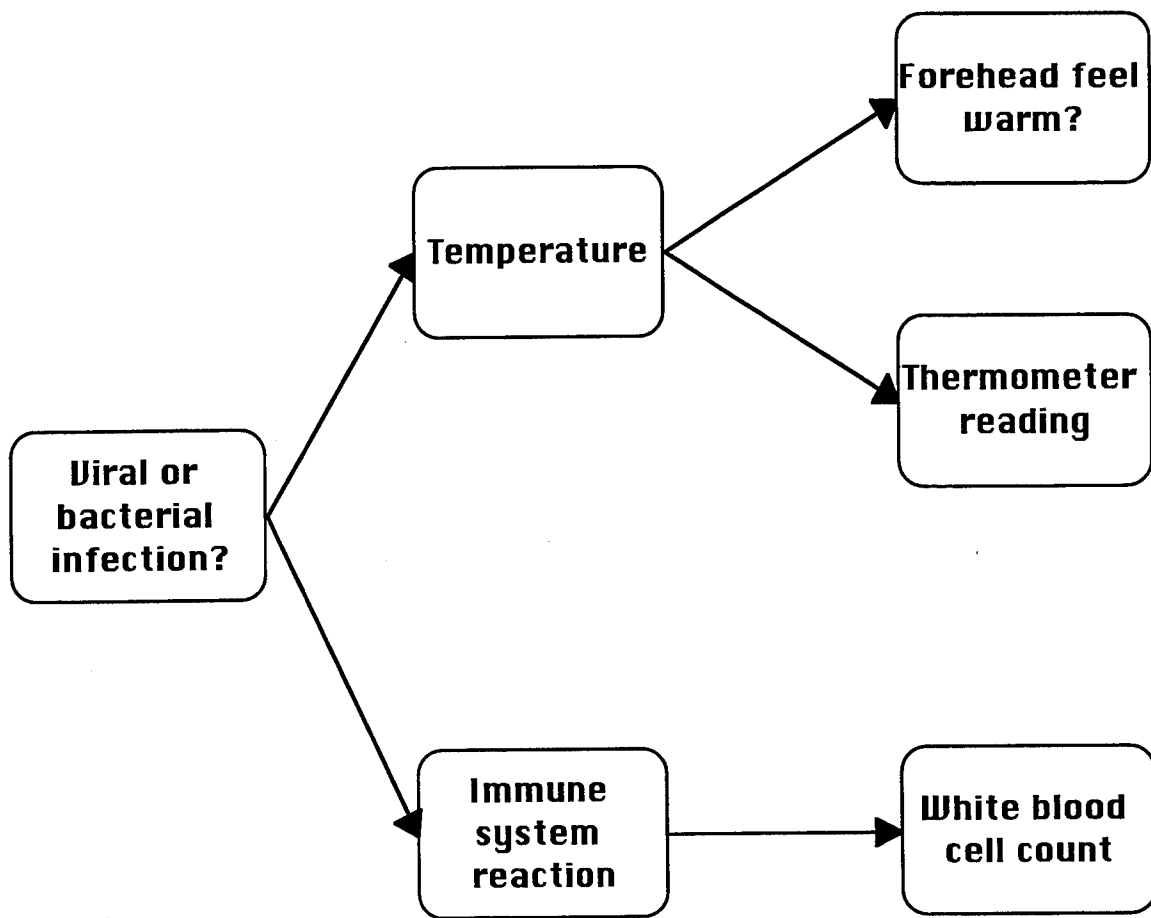


Figure 2

An Influence Diagram for a Simple Diagnostic Problem

Brophy 28 October 94

Distribution List

Dr Terry Ackerman  
Educational Psychology  
260C Education Bldg  
University of Illinois  
Champaign IL 61801

Dr Robert L Albright  
Educational Testing Service  
16-C Rosedale Road  
Princeton, NJ 08541

Dr Terry Allard  
Code 3422  
Office of Naval Research  
800 N Quincy St  
Arlington VA 22217-5660

Dr Nancy Allen  
Educational Testing Service  
Mail Stop 02-T  
Princeton NJ 08541

Dr Phipps Arabie  
Graduate School of Management  
Rutgers University  
92 New Street  
Newark NJ 07102-1895

Dr Isaac I Bejar  
Educational Testing Service  
Mail Stop 11-R  
Princeton NJ 08541

Dr William O Berry  
Director  
Life and Environmental Sciences  
AFOSR/NL N1  
Bldg 410  
Bolling AFB DC 20332-6448

Dr Thomas G Bever  
Department of Psychology  
University of Rochester  
River Station  
Rochester NY 14627

Dr Menucha Birenbaum  
School of Education  
Tel Aviv University  
Ramat-Aviv 69978 ISRAEL

Dr Bruce Bloxom  
Defense Manpower Data Center  
99 Pacific St  
Suite 155A  
Monterey CA 93943-3231

Dr Gwyneth Boodoo  
Educational Testing Service  
Mail Stop 03-T  
Princeton NJ 08541

Dr Richard L Branch  
HQ USMEPCOM/MEPCT  
2500 Green Bay Road  
North Chicago IL 60064

Dr Robert Brennan  
American College Testing  
2201 North Dodge Street  
PO Box 168  
Iowa City IA 52243

Dr David V Budescu  
Department of Psychology  
University of Haifa  
Mount Carmel Haifa 31999  
ISRAEL

Dr Gregory Candell  
CTB/MacMillan/McGraw-Hill  
2500 Garden Road  
Monterey CA 93940

Dr Susan Chipman  
Cognitive Science Program  
Office of Naval Research  
800 North Quincy Street  
Code 3422  
Arlington VA 22217-5660

Dr Raymond E Christal  
UES LAMP Science Advisor  
AL/HRMIL  
Brooks AFB TX 78235

Dr Norman Cliff  
Department of Psychology  
University of Southern California  
Los Angeles CA 90089-1061

Dr Nancy S Cole  
Educational Testing Service  
14-C Rosedale Road  
Princeton NJ 08541

Commanding Officer  
Naval Research Laboratory  
Code 4827  
Washington DC 20375-5000

Dr John M Cornwell  
Department of Psychology  
I/O Psychology Program  
Tulane University  
New Orleans LA 70118

Dr Linda Curran  
Defense Manpower Data Center  
Suite 400  
1600 Wilson Blvd  
Rosslyn VA 22209

Professor Clément Dassa  
Faculté des sciences de l'éducation  
Département d'études en éducation  
et d'administration de l'éducation  
CP 6128 succursale A  
Montréal Québec  
CANADA H3C 3J7

Dr Timothy Davey  
American College Testing  
2201 North Dodge Street  
PO Box 168  
Iowa City IA 52243

Dr Charles E Davis  
Educational Testing Service  
Mail Stop 16-T  
Princeton NJ 08541

Dr Ralph J DeAyala  
Meas Stat and Eval  
Benjamin Bldg Room 1230F  
University of Maryland  
College Park MD 20742

Director  
Life Sciences  
Code 3420  
Office of Naval Research  
Arlington VA 22217-5660

Hei-Ki Dong  
BELLCORE  
6 Corporate Place  
RM: PYA-1K207  
PO Box 1320  
Piscataway NJ 08855-1320

Dr Neil Dorans  
Educational Testing Service  
Mail Stop 07-E  
Princeton NJ 08541

Dr Fritz Drasgow  
University of Illinois  
Department of Psychology  
603 E Daniel Street  
Champaign IL 61820

Defense Tech Information Center  
Cameron Station Bldg 5  
Alexandria VA 22314  
(2 Copies)

Dr Richard Duran  
Graduate School of Education  
University of California  
Santa Barbara CA 93106

Dr Susan Embretson  
University of Kansas  
Psychology Department  
426 Fraser  
Lawrence KS 66045

Dr George Engelhard Jr  
Division of Educational Studies  
Emory University  
210 Fishburne Bldg  
Atlanta GA 30322

Dr Marshall J Farr  
Farr-Sight Co  
2520 North Vernon Street  
Arlington VA 22207

Dr Leonard Feldt  
Lindquist Center for Measurement  
University of Iowa  
Iowa City IA 52242

Dr Richard L Ferguson  
American College Testing  
2201 North Dodge Street  
PO Box 168  
Iowa City IA 52243

Dr Gerhard Fischer  
Liebiggasse 5  
A 1010 Vienna  
AUSTRIA

Dr Myron Fischl  
US Army Headquarters  
DAPE-HR  
The Pentagon  
Washington DC 20310-0300

Mr Paul Foley  
Navy Personnel R&D Center  
San Diego CA 92152-6800

Chair  
Department of Computer Science  
George Mason University  
Fairfax VA 22030

Dr Robert D Gibbons  
University of Illinois at Chicago  
NPI 909A M/C 913  
912 South Wood Street  
Chicago IL 60612

Dr Janice Gifford  
University of Massachusetts  
School of Education  
Amherst MA 01003

Dr Robert Glaser  
Learning Res & Development Cntr  
University of Pittsburgh  
3939 O'Hara Street  
Pittsburgh PA 15260

Dr Susan R Goldman  
Peabody College  
Box 45  
Vanderbilt University  
Nashville TN 37203

Dr Timothy Goldsmith  
Department of Psychology  
University of New Mexico  
Albuquerque NM 87131

Dr Sherrie Gott  
AFHRL/MOMJ  
Brooks AFB TX 78235-5601

Dr Bert Green  
Johns Hopkins University  
Department of Psychology  
Charles & 34th Street  
Baltimore MD 21218

Professor Edward Haertel  
School of Education  
Stanford University  
Stanford CA 94305-3096

Dr Ronald K Hambleton  
University of Massachusetts  
Lab of Psychom & Eval Res  
Hills South Room 152  
Amherst MA 01003

Dr Delwyn Harnisch  
University of Illinois  
51 Gerty Drive  
Champaign IL 61820

Dr Patrick R Harrison  
Computer Science Department  
US Naval Academy  
Annapolis MD 21402-5002

Ms Rebecca Hetter  
Navy Personnel R&D Center  
Code 13  
San Diego CA 92152-6800

Dr Thomas M Hirsch  
American College Testing  
2201 North Dodge Street  
PO Box 168  
Iowa City IA 52243

Professor Paul W Holland  
Div of Educ Psych & Quant  
Methods Prog  
Graduate School of Education  
4511 Tolman Hall  
University of California-Berkeley  
Berkeley CA 94720

Professor Lutz F Homke  
Institut für Psychologie  
RWTH Aachen  
Jaegerstrasse 17/19  
D-5100 Aachen  
WEST GERMANY

Ms Julia S Hough  
Cambridge University Press  
40 West 20th Street  
New York NY 10011

Dr William Howell  
Chief Scientist  
AFHRL/CA  
Brooks AFB TX 78235-5601

Dr Huynh Huynh  
College of Education  
University of South Carolina  
Columbia SC 29208

Dr Martin J Ippel  
Center for the Study of  
Education and Instruction  
Leiden University  
PO Box 9555  
2300 RB Leiden  
THE NETHERLANDS

Dr Robert Jannarone  
Elec and Computer Eng Dept  
University of South Carolina  
Columbia SC 29208

Dr Kumar Joag-dev  
University of Illinois  
Department of Statistics  
101 Illini Hall  
725 South Wright Street  
Champaign IL 61820

Professor Douglas H Jones  
Grad Sch of Management  
Rutgers The State University NJ  
Newark NJ 07102

Dr Brian Junker  
Carnegie-Mellon University  
Department of Statistics  
Pittsburgh PA 15213

Dr Marcel Just  
Carnegie-Mellon University  
Department of Psychology  
Schenley Park  
Pittsburgh PA 15213

Dr J L Kaiwi  
Code 442/JK  
Naval Ocean Systems Center  
San Diego CA 92152-5000

Dr Michael Kaplan  
Office of Basic Research  
US Army Research Institute  
5001 Eisenhower Avenue  
Alexandria VA 22333-5600

Dr Jeremy Kilpatrick  
Dept of Mathematics Education  
105 Aderhold Hall  
University of Georgia  
Athens GA 30602

Ms Hae-Rim Kim  
University of Illinois  
Department of Statistics  
101 Illini Hall  
725 South Wright Street  
Champaign IL 61820

Dr. Jwa-keun Kim  
Department of Psychology  
Middle Tennessee State University  
Murfreesboro TN 37132

Dr Sung-Hoon Kim  
KEDI  
92-6 Umyeon-Dong  
Seocho-Gu  
Seoul  
SOUTH KOREA

Dr G Gage Kingsbury  
Portland Public Schools  
Res & Eval Department  
501 North Dixon Street  
PO Box 3107  
Portland OR 97209-3107

Dr William Koch  
Box 7246  
Meas & Eval Center  
University of Texas-Austin  
Austin TX 78703

Dr James Kraatz  
Computer-based Education  
Research Laboratory  
University of Illinois  
Urbana IL 61801

Dr Patrick Kyllonen  
AFHRL/MOEL  
Brooks AFB TX 78235

Ms Carolyn Laney  
1515 Spencerville Rod  
Spencerville MD 20868

Richard Lanterman  
Commandant (G-PWP)  
US Coast Guard  
2100 Second Street SW  
Washington DC 20593-0001

Dr Michael Levine  
Educational Psychology  
210 Education Building  
1310 South Sixth Street  
Univ of IL at Urbana-Champaign  
Champaign IL 61820-6990

Dr Charles Lewis  
Educational Testing Service  
Mail Stop 03-T  
Princeton NJ 08541-0001

Mr Hsin-hung Li  
University of Illinois  
Department of Statistics  
101 Illini Hall  
725 South Wright Street  
Champaign IL 61820

Library  
Naval Training Systems Center  
12350 Research Parkway  
Orlando FL 32826-3224

Dr Marcia C Linn  
Graduate School of Education  
EMST  
Tolman Hall  
University of California  
Berkeley CA 94720

Dr Robert L Linn  
Campus Box 249  
University of Colorado  
Boulder CO 80309-0249

Dr Richard Luecht  
American College Testing  
2201 North Dodge Street  
PO Box 168  
Iowa City IA 52243

Dr. George Macready  
Professor  
Dept of Meas., Stat. & Eval.  
College of Education  
Room 1230C, Benjamin Bldg.  
University of Maryland  
College Park MD 20742

Dr Evans Mandes  
George Mason University  
4400 University Drive  
Fairfax VA 22030

Dr Paul Mayberry  
Center for Naval Analysis  
4401 Ford Avenue  
PO Box 16268  
Alexandria VA 22302-0268

Dr James R McBride  
HumRRO  
6430 Elmhurst Drive  
San Diego CA 92120

Mr Christopher McCusker  
University of Illinois  
Department of Psychology  
603 E Daniel Street  
Champaign IL 61820

Dr Joseph McLachlan  
Navy Pers Res & Dev Cntr  
Code 14  
San Diego CA 92152-6800

Alan Mead  
c/o Dr Michael Levine  
Educational Psychology  
210 Education Bldg  
University of Illinois  
Champaign IL 61801

Dr Timothy Miller  
American College Testing  
2201 North Dodge Street  
PO Box 168  
Iowa City IA 52243

Dr Robert Mislevy  
Educational Testing Service  
Mail Stop 03-T  
Princeton NJ 08541

Dr Ivo Molenaar  
Faculteit Sociale Wetenschappen  
Rijksuniversiteit Groningen  
Grote Kruisstraat 2/1  
9712 TS Groningen  
The NETHERLANDS

Dr Eiji Muraki  
Educational Testing Service  
Mail Stop 02-T  
Princeton NJ 08541

Dr Ratna Nandakumar  
Educational Studies  
Willard Hall Room 213E  
University of Delaware  
Newark DE 19716

Acad Prog & Research Branch  
Naval Tech Training Command  
Code N-62  
NAS Memphis (75)  
Millington TN 30854

Dr W Alan Nicewander  
American College Testing  
2201 North Dodge Street  
PO Box 168  
Iowa City IA 52243

Head  
Personnel Systems Department  
NPRDC (Code 12)  
San Diego CA 92152-6800

Director  
Training Systems Department  
NPRDC (Code 14)  
San Diego CA 92152-6800

Library NPRDC  
Code 041  
San Diego CA 92152-6800

Librarian  
Naval Cntr for Applied Research  
in Artificial Intelligence  
Naval Research Laboratory  
Code 5510  
Washington DC 20375-5000

Office of Naval Research  
Code 3422  
800 N Quincy Street  
Arlington VA 22217-5660  
(6 Copies)

ONR Resident Representative  
New York City  
33 Third Avenue - Lower Level  
New York NY 10003-9998

Special Asst for Res Management  
Chief of Naval Personnel  
(PERS-O1JT)  
Department of the Navy  
Washington DC 20350-2000

Dr Judith Orasanu  
NASA Ames Research Center  
Mail Stop 239-1  
Moffett Field CA 94035

Dr Peter J Pashley  
Law School Admission Services  
PO Box 40  
Newtown PA 18940-0040

Wayne M Patience  
American Council on Education  
GED Testing Service Suite 20  
One Dupont Circle NW  
Washington DC 20036

Dept of Administrative Sciences  
Code 54  
Naval Postgraduate School  
Monterey CA 93943-5026

Dr Peter Pirolli  
School of Education  
University of California  
Berkeley CA 94720

Dr Mark D Reckase  
American College Testing  
2201 North Dodge Street  
PO Box 168  
Iowa City IA 52243

Mr Steve Reise  
Department of Psychology  
University of California  
Riverside CA 92521

Mr Louis Roussos  
University of Illinois  
Department of Statistics  
101 Illini Hall  
725 South Wright Street  
Champaign IL 61820

Dr Donald Rubin  
Statistics Department  
Science Center Room 608  
1 Oxford Street  
Harvard University  
Cambridge MA 02138

Dr Fumiko Samejima  
Department of Psychology  
University of Tennessee  
310B Austin Peay Bldg  
Knoxville TN 37966-0900

Dr Mary Schratz  
4100 Parkside  
Carlsbad CA 92008

Mr Robert Semmes  
N218 Elliott Hall  
Department of Psychology  
University of Minnesota  
Minneapolis MN 55455-0344

Dr Valerie L Shalin  
Dept of Industrial Engineering  
State University of New York  
342 Lawrence D Bell Hall  
Buffalo NY 14260

Mr Richard J Shavelson  
Graduate School of Education  
University of California  
Santa Barbara CA 93106

Kathleen Sheehan  
Educational Testing Service  
Mail Stop 03-T  
Princeton NJ 08541

Dr Kazuo Shigemasu  
7-9-24 Kugenuma-Kaigan  
Fujisawa 251  
JAPAN

Dr Randall Shumaker  
Naval Research Laboratory  
Code 5500  
4555 Overlook Avenue SW  
Washington DC 20375-5000

Dr Judy Spray  
American College Testing  
2201 North Dodge Street  
PO Box 168  
Iowa City IA 52243

Dr Martha Stocking  
Educational Testing Service  
Mail Stop 03-T  
Princeton NJ 08541

Dr William Stout  
University of Illinois  
Department of Statistics  
101 Illini Hall  
725 South Wright St  
Champaign IL 61820

Dr Kikumi Tatsuoka  
Educational Testing Service  
Mail Stop 03-T  
Princeton NJ 08541

Dr David Thissen  
Psychometric Laboratory  
CB# 3270 Davie Hall  
University of North Carolina  
Chapel Hill NC 27599-3270

Mr Thomas J Thomas  
Federal Express Corporation  
Human Resource Development  
3035 Director Row Suite 501  
Memphis TN 38131

Mr Gary Thomasson  
University of Illinois  
Educational Psychology  
Champaign IL 61820

Dr Howard Wainer  
Educational Testing Service  
15-T Rosedale Road  
Princeton NJ 08541

Elizabeth Wald  
Office of Naval Technology  
Code 227  
800 North Quincy Street  
Arlington VA 22217-5000

Dr Michael T Waller  
Univ of Wisconsin-Milwaukee  
Educ Psychology Department  
Box 413  
Milwaukee WI 53201

Dr Ming-Mei Wang  
Educational Testing Service  
Mail Stop 03-T  
Princeton NJ 08541

Dr Thomas A Warm  
FAA Academy  
PO Box 25082  
Oklahoma City OK 73125

Dr David J Weiss  
N660 Elliott Hall  
University of Minnesota  
75 E River Road  
Minneapolis MN 55455-0344

Dr Douglas Wetzel  
Code 15  
Navy Personnel R&D Center  
San Diego CA 92152-6800

German Military Representative  
Personalstammamt  
Koelner Str 262  
D-5000 Koeln 90  
WEST GERMANY

Dr David Wiley  
Sch of Educ and Social Policy  
Northwestern University  
Evanston IL 60208

Dr Bruce Williams  
Dept of Educational Psychology  
University of Illinois  
Urbana IL 61801

Dr Mark Wilson  
School of Education  
University of California  
Berkeley CA 94720

Dr Eugene Winograd  
Department of Psychology  
Emory University  
Atlanta GA 30322

Dr Martin F Wiskoff  
PERSEREC  
99 Pacific Street  
Suite 4556  
Monterey CA 93940

Mr John H Wolfe  
Navy Personnel R&D Center  
San Diego CA 92152-6800

Dr Kentaro Yamamoto  
Educational Testing Service  
Mail Stop 03-T  
Princeton NJ 08541

Duanli Yan  
Educational Testing Service  
Mail Stop 03-T  
Princeton NJ 08541

Dr Wendy Yen  
CTB/McGraw Hill  
Del Monte Research Park  
Monterey CA 93940

Dr Joseph L Young  
National Science Foundation  
Room 320  
1800 G Street NW  
Washington DC 20550